ADVANCED PROTEOMIC CHARACTERIZATION OF THE 26S PROTEASOME IN ARABIDOPSIS REVEALS INSIGHTS INTO COMPOSITION AND ASSEMBLY

by

David C. Gemperline

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Genetics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2016

Date of final oral examination: TBD

The dissertation is approved by the following members of the Final Oral Committee:
    Richard D. Vierstra, Professor, Genetics
    Richard Amasino, Professor, Genetics
    Josh Coon, Professor, Chemistry
    Donna Fernandez, Professor, Botany
    Patrick Masson, Professor, Genetics

To Erin, my wife.

ACKNOWLEDGMENTS

Science doesn't purvey absolute truth. Science is a mechanism, a way of trying to improve your knowledge of nature. It's a system for testing your thoughts against the universe and seeing whether they match. This works not just for the ordinary aspects of science, but for all of life.

— ISAAC ASIMOV (1988)

Acknowledgements go here.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

API	Application programming interface
BCA	Bicinchoninic acid protein assay
BLAST	Basic local alignment search tool
C#	C sharp, a programming language
Da	Dalton, the atomic mass unit
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ESI	Electrospray ionization
E-value	Expectation value
FASTA	A format for storing protein sequences
FDR	False discovery rate
GUI	Graphical user interface
HCD	Higher-energy collisional dissociation
HPLC	High-performance liquid chromatography
LC	Liquid chromatography
m	Mass
min	Minute
MS	Mass spectrometry

MS^1	Survey mass analysis
MS/MS	Tandem mass spectrometry
NCE	Normalized collision energy
nLC	Nanoflow liquid chromatography
ppm	Part per million
PSM	Peptide-spectrum match
PTM	Post-translational modification
s	Second
SILAC	Stable isotope labeling by amino acids in cell culture
S/N	Signal-to-noise ratio
TMT	Tandem mass tag

ADVANCED PROTEOMIC CHARACTERIZATION OF THE 26S PROTEASOME IN ARABIDOPSIS REVEALS INSIGHTS INTO COMPOSITION AND ASSEMBLY

David C. Gemperline

Under the supervision of Professor Richard D. Vierstra

At the University of Wisconsin-Madison

Abstract

The 26S proteasome is the central proteolytic effector in the ubiquitin system that is responsible for degrading numerous regulators following their selective ubiquitylation. While much is known about the construction of the yeast and mammal particles, little is known about the pathways used to assemble the plant particle. One challenge is that the known yeast chaperones appear sufficiently diverged to preclude high-confidence identification of their plant counterparts by genomic searches. Here, we used in-depth mass spectrometric analysis of Arabidopsis 26S proteasomes, which were affinity purified from seedlings under conditions that promote the accumulation of assembly intermediates, to identify a large collection of interacting proteins that associate with either the core protease (CP) or regulatory particle (RP). Sequence comparisons, Y2H and BiFC studies revealed

that some are likely assembly chaperones, with several CP factors harboring the signature C-terminal HbYX motif that allows their association with the α-subunit ring. Several of the RP-specific factors appear to be orthologs of the chaperones Nas2, Nas6, Hsm3 and Ecm29. Whereas yeast assembles only a single particle type, mammals can assemble alternate proteasomes by replacing individual subunits with distinct isoforms (e.g., immunoproteasomes). In plants, most 26S proteasome subunits are encoded by paralogous genes with sufficient divergence to suggest that plants also accumulate a collection of particles. However, proteomic analysis of proteasomes selectively enriched using paralog-specific tags strongly imply that although plants possess this genetic diversity, the incorporation of these paralogs appears random, and is mainly influenced by the differential expression of the corresponding genes. Taken together, these proteomic studies provide the first insights into plant proteasome assembly and diversity, and identify factors that build the CP and RP subcomplexes and finally the 26S holo-particle.

Richard D. Vierstra

ABSTRACT

Abstract

The 26S proteasome is the central proteolytic effector in the ubiquitin system that is responsible for degrading numerous regulators following their selective ubiquitylation. While much is known about the construction of the yeast and mammal particles, little is known about the pathways used to assemble the plant particle. One challenge is that the known yeast chaperones appear sufficiently diverged to preclude high-confidence identification of their plant counterparts by genomic searches. Here, we used in-depth mass spectrometric analysis of Arabidopsis 26S proteasomes, which were affinity purified from seedlings under conditions that promote the accumulation of assembly intermediates, to identify a large collection of interacting proteins that associate with either the core protease (CP) or regulatory particle (RP). Sequence comparisons, Y2H and BiFC studies revealed that some are likely assembly chaperones, with several CP factors harboring the signature C-terminal HbYX motif that allows their association with the α-subunit ring. Several of the RP-specific factors appear to be orthologs of the chaperones Nas2, Nas6, Hsm3 and Ecm29. Whereas yeast assembles only a single particle type, mammals can assemble alternate proteasomes by replacing individual subunits with distinct isoforms (e.g., immunoproteasomes). In plants, most 26S proteasome

subunits are encoded by paralogous genes with sufficient divergence to suggest that plants also accumulate a collection of particles. However, proteomic analysis of proteasomes selectively enriched using paralog-specific tags strongly imply that although plants possess this genetic diversity, the incorporation of these paralogs appears random, and is mainly influenced by the differential expression of the corresponding genes. Taken together, these proteomic studies provide the first insights into plant proteasome assembly and diversity, and identify factors that build the CP and RP subcomplexes and finally the 26S holo-particle.

Chapter 1

THE UBIQUITIN 26S PROTEASOME SYSTEM

1.1. Ubiquitin Conjugating Machinery

1.1.1. E1s

1.1.2. E2s

1.1.3. E3s

1.1.4. DUBS

1.2. The 26S Proteasome

1.2.1. The 20S Core Protease

1.2.2. The 19S Regulatory Particle

1.2.2.1. Regulatory Particle Base

1.2.2.2. Regulatory Particle Lid

1.3.   Proteasome Expression

1.4.   Proteasome Assembly

1.5.   Proteasome Post-Translational Modification

1.6.   Proteasome Degredation

1.7.   Proteasome Interacting Proteins

Chapter 2

MORPHEUS SPECTRAL COUNTER: A COMPUTATIONAL TOOL FOR LABEL-FREE QUANTITATIVE MASS SPECTROMETRY USING THE MORPHEUS SEARCH ENGINE

2.1. Summary

Label-free quantitative MS based on the Normalized Spectral Abundance Factor (NSAF) has emerged as a straightforward and robust method to determine the relative abundance of individual proteins within complex mixtures. Here, we present Morpheus Spectral Counter (MSpC) as the first computational tool that directly calculates NSAF values from output obtained from Morpheus, a fast, open-source, peptide-MS/MS matching engine compatible with high-resolution accurate-mass instruments. NSAF has distinct advantages over other MS-based quantification methods, including a higher dynamic range as compared to isobaric tags, no requirement to align and re-extract MS1 peaks, and increased speed. MSpC features an easy to use graphic user interface that additionally calculates both distributed and unique NSAF values to permit analyses of both protein families and isoforms/proteoforms. MSpC determinations of protein concentration were linear over

several orders of magnitude based on the analysis of several high-mass accuracy datasets either obtained from PRIDE or generated with total cell extracts spiked with purified *Arabidopsis* 20S proteasomes. The MSpC software was developed in C# and is open sourced under a permissive license with the code made available at `http://dcgemperline.github.io/Morpheus_SpC/`.

2.2. Main Text

Quantification of individual polypeptides within complex mixtures by MS is an extremely useful tool to understand proteomic changes in organisms during growth and development, and after environmental perturbation (Wong and Cagney, 2010). While a number of MS/MS strategies have been developed to measure protein abundance, including Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC), labeling with isobaric tags, and Absolute Quantification of proteins (AQUA) (Gerber et al., 2003; Ong et al., 2002; Ross et al., 2004; Thompson et al., 2003), label-free quantification (LFQ) have become increasingly popular given their simplicity and low cost (Wong and Cagney, 2010; Zhang et al., 2006). One LFQ strategy infers abundance from the number of observed peptide spectra matches (PSMs). For these PSM-based approaches, changes in protein abundance can be generated artifactually when total PSMs differ among samples and because longer proteins

tend to produce more raw counts. For these reasons normalizing for both protein length and total PSMs is paramount. While this adjustment can be made in a number of ways; one of the most straight forward methods is to use Normalized Spectral Abundance Factor (NSAF), a length- and count-normalized measure for each protein (Zybailov et al., 2006). Further improvements to the NSAF algorithm have been made by accounting for shared peptides in distributed NSAF (dNSAF), which distributes common PSMs among a family of isoforms/proteoforms based on the number of distinct PSMs observed for each isoform/proteoform, and unique NSAF (uNSAF), which ignores shared PSMs and only assigns distinct PSMs to each specific isoform/proteoform (Zhang et al., 2010).

The Morpheus MS search engine was recently designed for high-resolution, accurate-mass data obtained from Orbitrap-based instruments to provide faster matching of spectra to peptides (Wenger and Coon, 2013). Unfortunately, no downstream automated tools are available to facilitate LFQ analysis, which can be quite challenging, if not impossible, to complete manually when accounting for shared peptides. To overcome this bottleneck, we developed Morpheus Spectral Counter (MSpC) as the first LFQ computational tool that integrates directly with Morpheus to calculate NSAF, dNSAF, uNSAF, and corrected PSM (Fermin et al., 2011) values in complex protein samples. MSpC is fully automated, and only requires a
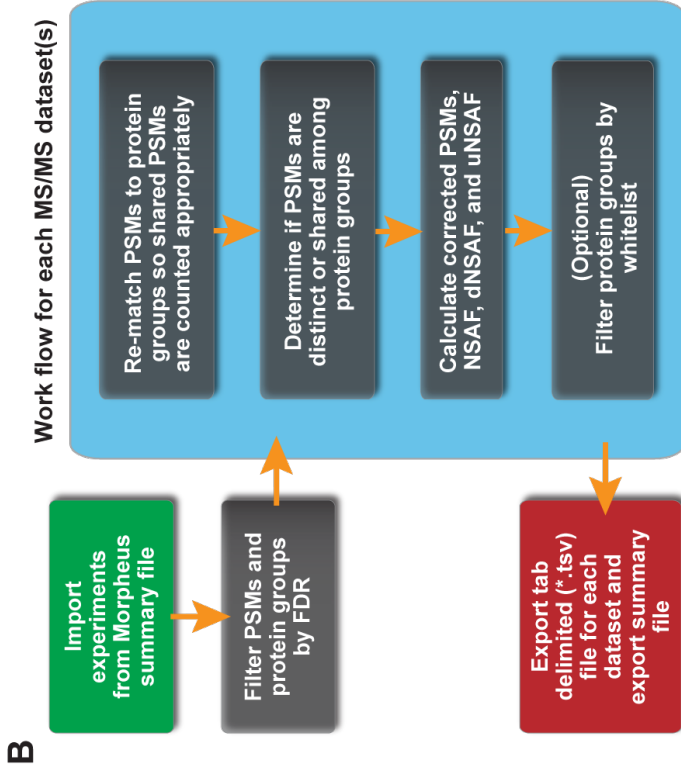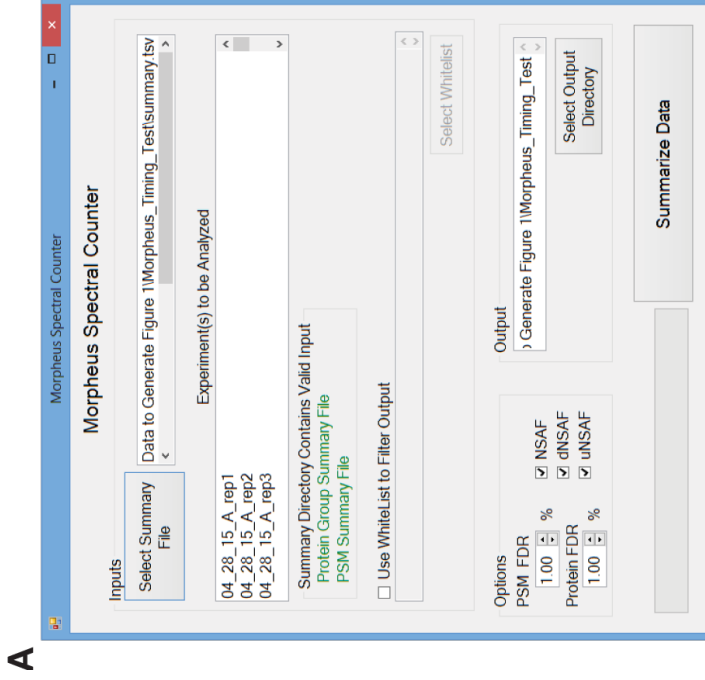
Figure 2.1: MSpC Graphic User Interface (GUI) and software flow chart. (A) Screenshot of the GUI. The input requires the user to select a Morpheus search summary file containing experiments to be analyzed. The user can optionally select a whitelist to filter output, and select an output directory. Additional options can set peptide and protein FDR cutoffs, and method of quantification for output, including Normalized Spectral Abundance Factor (NSAF), distributed NSAF (dNSAF), and unique NSAF (uNSAF). A progress bar highlights completion of the analysis. (B) Data analysis flow chart. Experiments and groups of experiments to be analyzed are imported through the Morpheus summary file. PSMs and protein groups are filtered at the specified FDR cutoff with a default of 1%.

Morpheus search summary file (summary.tsv) as input. The user interface (see Figure 2.1 (A)) allows one to select the summary file and displays the raw MS/MS files that will be analyzed by MSpC. Due to shared peptides being attributed to only one instance of a protein group in Morpheus's PSM file, PSMs are re-matched to all possible protein groups. PSMs are then cataloged as shared or as unique (distinctly matching one protein group) to generate NSAF, dNSAF, and uNSAF outputs. Finally, the output can be filtered for proteins of interest by specifying a comma delimited file containing unique identifiers and descriptions. Some important features of MSpC are its ability to handle fractionation experiments as input, and the ability to whitelist proteins of interest in the output by specifying a csv file (see Tutorial 2.5). Options exist to specify global PSM and protein group FDR rates (thus avoiding increased FDRs when one analyzes many experiments at once), to output NSAF, dNSAF, and uNSAF values, to require a minimum number of unique peptides to quantify a protein, and to specify an output directory. A progress bar indicates completion of the analysis by MSpC. To validate the accuracy of MSpC, we analyzed two MS/MS datasets available in PRIDE that were previously generated by high-energy collision-induced dissociation using Thermo Q-Exactive Orbitrap instruments.

Here, Xenopus egg (see top, Figure 2.2) and embryo (bottom, Figure 2.2) extracts
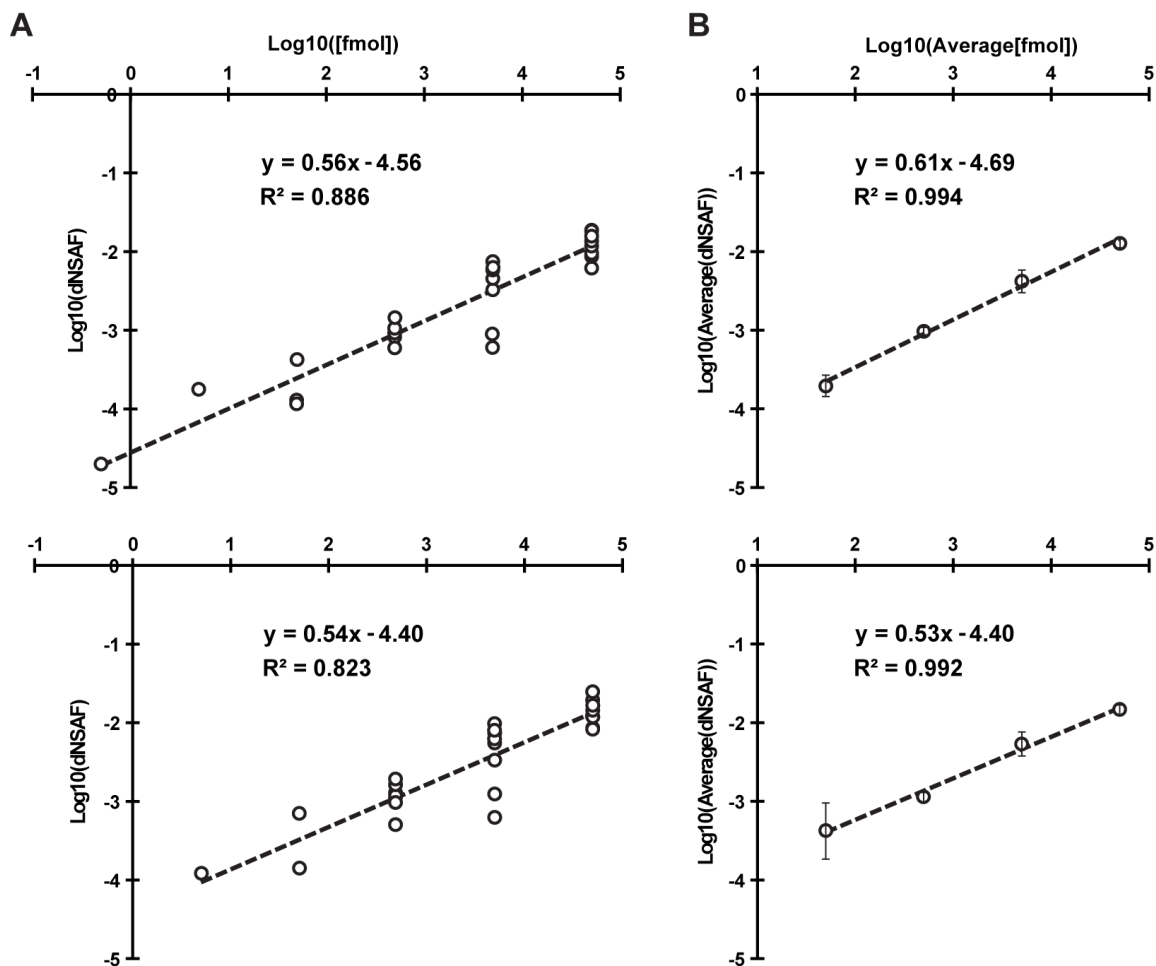
Figure 2.2: Confirmation of MSpC accuracy by analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2). The array of UPS2 standards were spiked into Xenopus laevis egg **(**Top**)** and embryo **(**Bottom**)** extracts at a range of concentrations. Following MS/MS analysis, dNSAF values for each protein were determined by Morpheus and MSpC. **(**A**)** A log-log plot of dNSAF versus concentration for each UPS2 protein detected across each fmol range. **(**B**)** A log-log plot of average dNSAF vs average concentration of each group of UPS2 proteins at each fmol range: (50, 500, 5000, and 50,000 fmol).

were spiked at a 4:1 ratio with the Universal Proteome Standard 2 (UPS2), a mix of 48 purified proteins at defined molar ratios of 0.5, 5, 50, 500, 5000, and 50,000, with each ratio containing a different set of 8 of the 48 proteins. As shown in Figure 2.2 A, when the Morpheus/MSpC pipeline was used to calculate the average dNSAF value for each UPS2 protein, requiring only a single unique peptide to quantify, strong linear correlations (R^2 = 0.886 and 0.823) were obtained across a 1,000 fold change in abundance (50 fmol to 50,000 fmol). In fact, the R^2 values were similar to those obtained by others with PSM-based LFQ methods (Cox et al., 2014; Tu et al., 2014). This linear correlation was further strengthened when the dNSAF values were averaged for all UPS2 proteins within each of the concentration groups, with R^2 values of 0.994 and 0.992 for the egg and embryo datasets, respectively (see Figure 2.2 B). Notably, the slope of the concentration series was significantly less than unity, showing that NSAF measurements are not appropriate for absolute quantification, which was expected given that NSAF is a relative value.

We also reprocessed the UPS2 dataset using the option of requiring a minimum of two unique peptides for quantification, which should improve stringency. This option provided only a minor improvement in overall linearity for the average UPS2 dNSAF values, but decreased linearity when each UPS2 protein was considered individually and removed some UPS2 proteins at low concentrations (compare

A

Log10([fmol])

$y = 0.56x - 4.58$
$R^2 = 0.865$

Log10(dNSAF)

B

Log10(Average[fmol])

$y = 0.61x - 4.71$
$R^2 = 0.995$

Log10(Average(dNSAF))

$y = 0.55x - 4.47$
$R^2 = 0.727$

Log10(dNSAF)

$y = 0.53x - 4.29$
$R^2 = 0.992$

Log10(Average(dNSAF))

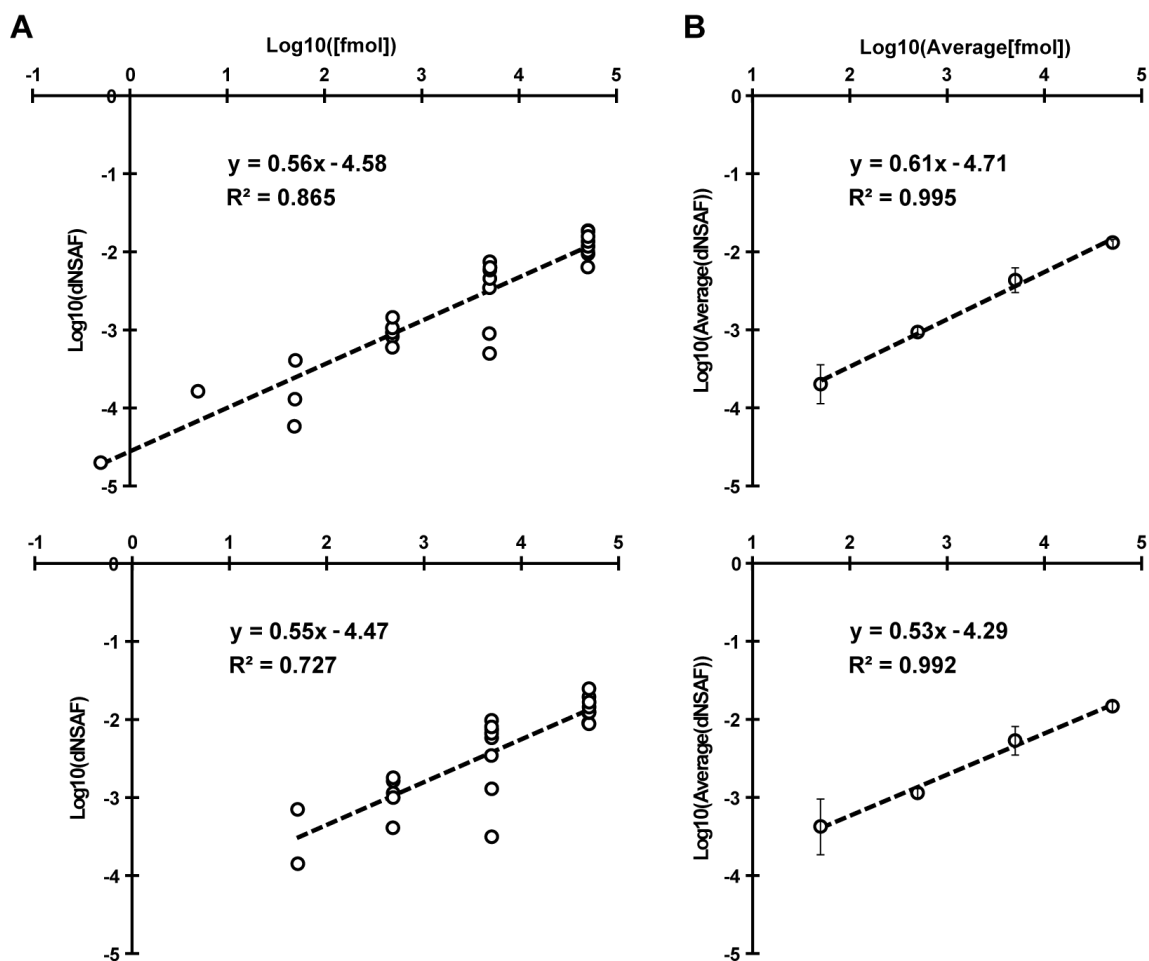**Figure 2.3: Re- Analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2).** The array of UPS2 standards were spiked into Xenopus laevis egg **(**Top) and embryo **(**Bottom) extracts at a range of concentrations. Following MS/MS analysis, dNSAF values for each protein were determined by Morpheus and MSpC with a change from Figure 2.2 in that two unique peptides were required to quantify a protein. **(**A) A log-log plot of dNSAF versus concentration for each UPS2 protein detected across each fmol range. **(**B) A log-log plot of average dNSAF vs average concentration of each group of UPS2 proteins at each fmol range: (50, 500, 5000, and 50,000 fmol).

Figure 2.3 A to Figure 2.2 A). Consequently, caution should be exercised when selecting this option even though it might provide a slight improvement in stringency (see Discussion on Requiring Two Peptide 2.4). To demonstrate the utility and accuracy of MSpC as applied to our work, we analyzed 20S proteasomes isolated from *Arabidopsis thaliana*. This particle contains multiple subunits assembled in stoichiometric amounts, with many subunits encoded by two paralogous genes of sufficient amino acid identity (typically >90% (Yang et al., 2004)) such that discrimination between paralogs can be challenging using LFQ approaches (Book et al., 2010). To simulate changes in 20S proteasome abundance, we added varying amounts of trypsinized proteasomes (0.05 μg to 3 μg) to a fixed amount of trypsinized *Escherichia coli* lysate (0.5 μg) to generate proteasome/lysate ratios of ~0.091, 0.167, 0.333, 0.500, 0.667, 0.750 0.800, 0.857. The digests were then subjected to MS/MS and the dNSAF value for each subunit along with the uNSAF value for individual isoforms were calculated by the Morpheus/MSpC pipeline (see Methods 2.3). The data from this experiment are deposited in PRIDE with ID PXD003002. As shown in Figure 2.4, MSpC provided an excellent determination for the overall abundance of 20S proteasomes within a complex mixture, along with a good reflection of the abundance of individual subunits and their isoforms.

When the dNSAF values for all subunits for the *Arabidopsis* 20S proteasome

[Figure 2.4 *caption follows on next page*]

A

B

$y = 0.875x + 0.0707$
$R^2 = 0.9915$

E
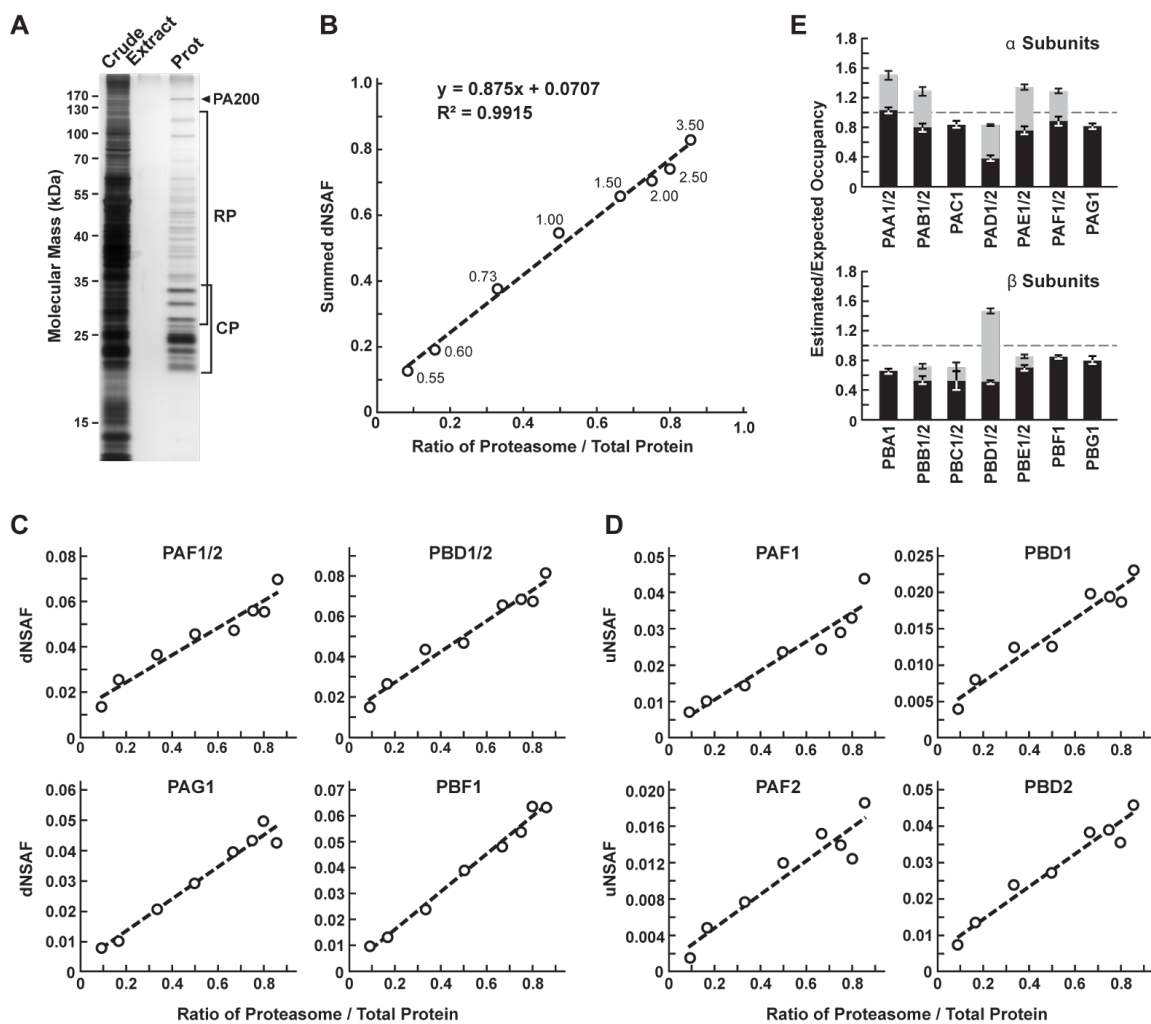
including their isoforms (representing 14 distinct subunits, 10 of which exist as

isoform pairs) were summed, a very close approximation of the dNSAF/actual

abundance was obtained (slope=0.875) with a very strong linear correlation (R^2

= 0.99) over a ~10-fold range in protein abundance. When each 20S proteasome

subunit was analyzed individually, a strong linear response was also obtained (R^2

> 0.90) for a majority of subunits (Figure 2.4 C and Table 2.1).

For example, reasonably accurate concentration plots were obtained for the

Figure 2.4 *(preceding page)*: **Confirmation of MSpC accuracy by analysis of MS/MS datasets generated with affinity purified *Arabidopsis* 20S proteasomes spiked into a total cell lysate from *E. coli*. Following MS/MS analysis, the dNSAF and uNSAF values for each subunit/isoform were determined by Morpheus and MSpC. (A)** A silver-stained SDS-PAGE gel of 20S proteasome samples affinity purified from 10-d-old *Arabidopsis* seedlings. The crude seedling extract(CE), sample buffer (SB), and affinity-purified 20S proteasome samples (Prot) are shown. **(B)** Quantification of trypsinized 20S proteasomes when mixed at varying ratios with trypsinized total protein lysates from *E. coli*. The spiked samples were subjected to MS/MS followed by data analysis with the Morpheus and MSpC. dNSAF values for each proteasome subunit were averaged across three technical replicates, then summed to obtain an estimate of abundance for the 20S proteasome, and plotted against their known ratios. The total protein load is listed at each point in ug. **(C and D)** dNSAF and uNSAF values determined from the data in panel B for individual subunits **(C)** and their isoforms **(D)** for several subunits of the 20S proteasome. **(E)** Quantification accuracy of the Morpheus/MSpC pipeline for determining the amount of each α and β subunit of the 20S proteasome. SIngle subunit isoforms are in black, wheras subunits having two isoforms are shown in black and grey to reflect the contributions of isoforms 1 and 2 respectively. Each bar represents the average of eight technical replicates (± SE). The dashed line represents the expected value of one assuming an equal stoichiometry of each subunit within the particle.

PAF (α6) and PBD (β4) subunits that are encoded by the PAF1/2 and PBF1/2 gene pairs, and for the PAG (α7) and PBF (β6) subunits that are encoded by single PAG1 and PBF1 genes (R^2 from 0.94 to 0.99). Even when we calculated uNSAF values for individual isoforms added to the *E. coli* lysate, strong linear responses were obtained (e.g., the PAF1/PAF2 and PBD1/PBD2 pairs) with robust correlations (R^2 from 0.89 to 0.95) (Figure fig:proteasomespike D). Taken together, MSpC worked well for relative LFQ analysis of a multi-subunit complex and its individual subunits and isoforms within a complex proteomic mixture.

The Morpheus/MSpC pipeline also allowed us to calculate the respective incorporation of each paralog in the complex (see Isoform Incorporation Methods 2.3.4). As shown in Figure 2.4E, these estimated/expected occupancies were close to unity for most subunits within both the α and β rings of the 20S proteasome. The only strong deviation was for PBD1/2 (β4), which had a greater dNSAF value relative to other β subunits across the experiments analyzed (see Table 2.1). The calculations for uNSAF values also estimated the relative proportion of each isoform within the complex for those subunits expressed from paralogous genes. The data obtained are similar to prior studies of the complex involving quantitative top-down proteomic analysis of purified proteasome samples using ultra violet-intrinsic fluorescence to quantify tyrosine-containing subunits (Russell et al., 2013). However, our MSpC

analysis provided a more complete picture as several subunit isoforms were difficult to quantify by fluorescence either because they lacked tryosine, or because their fluorescence peaks overlapped with those of other subunits/isoforms. Notably , the protein isoform ratios measured here agree well with the expression ratios for the paralogous genes (Book et al., 2010), suggesting that the protein isoform abundance generally reflects the relative transcriptional activity of the gene pair. We consistently estimated slightly more α ring subunits (PAA-PAG) versus β ring subunits (PBA-PBG) in the final MSpC calculations (Figure 2.4 E). This deviation could represent enhanced detection of α ring versus β ring subunits, or more likely that purification via the tagged α ring subunit PAG1 also isolated assembly intermediates comprised of only α ring subunits.

We compared the Morpheus and MSpC pipeline to the next most comparable open source, spectral-count-based LFQ pipeline, The Trans Proteomic Pipeline (TPP) (Deutsch et al., 2010) and ABACUS (Fermin et al., 2011) using our datasets generated with the 20S proteasome/*E. coli* lysate mixture (see Tables 2.1 and 2.2). Morpheus/MSpC slightly outperformed TPP/ABACUS by having a greater overall accuracy (average linearity of 0.88 compared to 0.84), and by having more subunits showing an R^2 linear correlation greater than 0.9 (14/23 subunits for MSpC versus and 11/23 for ABACUS).

Table 2.1: Table of dNSAF values for each 20S proteasome subunit generated by analyzing the proteasome spike in experiments with the Morpheus and MSpC pipeline. The top half of the table lists α1-7 (PAA-PAG) where 1 or 2 represent different isoforms) subunits, while the bottom half of the table lists β1-7 subunits (PBA-PBG where 1 or 2 represent different isoforms)

Ratio	0.091	0.167	0.333	0.500	0.667	0.750	0.800	0.857		Pearsons
PAA1	0.0102	0.0180	0.0275	0.0392	0.0505	0.0489	0.0467	0.0530		0.952
PAA2	0.0063	0.0072	0.0118	0.0121	0.0139	0.0218	0.0351	0.0195		0.652
PAB1	0.0091	0.0117	0.0209	0.0391	0.0377	0.0377	0.0268	0.0397		0.721
PAB2	0.0017	0.0062	0.0147	0.0223	0.0295	0.0277	0.0303	0.0250		0.894
PAC1	0.0076	0.0092	0.0210	0.0339	0.0373	0.0395	0.0546	0.0512		0.955
PAD1	0.0056	0.0070	0.0103	0.0151	0.0160	0.0147	0.0149	0.0147		0.826
PAD2	0.0039	0.0068	0.0141	0.0162	0.0200	0.0225	0.0255	0.0228		0.959
PAE1	0.0046	0.0095	0.0208	0.0352	0.0395	0.0356	0.0389	0.0490		0.928
PAE2	0.0063	0.0082	0.0184	0.0240	0.0269	0.0282	0.0266	0.0293		0.925
PAF1	0.0106	0.0173	0.0239	0.0303	0.0292	0.0379	0.0406	0.0492		0.920
PAF2	0.0032	0.0083	0.0128	0.0155	0.0182	0.0184	0.0152	0.0209		0.847
PAG1	0.0080	0.0100	0.0208	0.0291	0.0394	0.0431	0.0498	0.0423		0.966
Ratio	0.091	0.167	0.333	0.500	0.667	0.750	0.800	0.857		Pearsons
PBA1	0.0076	0.0104	0.0170	0.0207	0.0246	0.0288	0.0337	0.0449		0.898
PBB1	0.0066	0.0074	0.0191	0.0276	0.0223	0.0206	0.0172	0.0227		0.498
PBB2	0.0009	0.0016	0.0034	0.0035	0.0110	0.0109	0.0160	0.0161		0.891
PBC1	0.0000	0.0000	0.0162	0.0309	0.0373	0.0344	0.0295	0.0388		0.877
PBC2	0.0000	0.0000	0.0000	0.0096	0.0144	0.0147	0.0146	0.0190		0.931
PBD1	0.0053	0.0097	0.0150	0.0147	0.0221	0.0228	0.0231	0.0271		0.955
PBD2	0.0094	0.0162	0.0286	0.0318	0.0426	0.0457	0.0441	0.0538		0.969
PBE1	0.0070	0.0097	0.0156	0.0224	0.0323	0.0373	0.0334	0.0519		0.914
PBE2	0.0007	0.0003	0.0036	0.0069	0.0080	0.0105	0.0106	0.0133		0.973
PBF1	0.0093	0.0122	0.0205	0.0310	0.0368	0.0412	0.0473	0.0459		0.991
PBG1	0.0085	0.0165	0.0203	0.0323	0.0331	0.0345	0.0334	0.0442		0.912
									Average	0.885

In addition to this modest improvement, we note that the Morpheus/MSpC pipeline required significantly less intermediary steps, thus accelerating the data analysis. Some of the additional steps in TPP/ABACUS could be automated from the command-line, but it would likely be a challenge for the average user. Importantly, we found that the Morpheus/MSpC pipeline was faster. Timing tests using the proteasome/*E. coli* spike data generated here showed that the Morpheus/MSpC

Table 2.2: Table of adj_NSAF values for each 20S proteasome subunit (equivalent to dNSAF) generated by analyzing the proteasome spike in experiments with the TPP and ABACUS pipeline. The top half of the table lists α1-7 (PAA-PAG) where 1 or 2 represent different isoforms) subunits, while the bottom half of the table lists β1-7 subunits (PBA-PBG where 1 or 2 represent different isoforms)

Ratio	0.091	0.167	0.333	0.500	0.667	0.750	0.800	0.857		Pearsons
PAA1	125.16	166.81	245.61	355.60	491.76	494.18	504.75	554.97		0.988
PAA2	57.52	70.52	114.60	126.35	132.06	162.25	358.13	154.62		0.493
PAB1	96.13	134.98	207.90	328.16	362.21	416.46	303.60	397.39		0.865
PAB2	29.67	79.42	153.34	282.91	365.22	299.47	310.19	299.93		0.849
PAC1	95.27	112.45	214.24	376.66	389.05	361.39	512.73	488.62		0.925
PAD1	62.58	76.16	106.03	196.03	179.37	182.99	164.04	188.71		0.794
PAD2	39.88	68.33	145.34	178.43	190.84	246.20	243.28	240.01		0.954
PAE1	50.41	97.53	193.18	368.87	464.79	397.27	468.64	534.16		0.951
PAE2	71.38	80.07	184.85	255.98	299.41	290.16	304.77	348.32		0.955
PAF1	132.67	179.60	207.68	253.51	253.04	327.11	334.73	460.76		0.835
PAF2	0.00	0.00	59.62	70.93	111.83	80.29	63.08	65.86		0.606
PAG1	108.09	142.82	271.18	389.94	495.94	530.90	632.72	540.86		0.966
Ratio	**0.091**	**0.167**	**0.333**	**0.500**	**0.667**	**0.750**	**0.800**	**0.857**		**Pearsons**
PBA1	85.99	113.01	208.38	233.44	273.50	327.23	417.12	560.62		0.851
PBB1	47.77	80.66	195.35	324.53	303.26	220.36	200.51	216.77		0.440
PBB2	17.11	4.54	0.00	0.00	21.05	102.51	144.30	187.86		0.618
PBC1	27.19	40.65	165.46	345.14	364.19	386.59	338.47	373.16		0.885
PBC2	10.64	40.65	0.00	55.58	145.43	158.90	159.74	228.05		0.845
PBD1	50.14	94.46	168.58	167.50	244.25	248.17	252.65	319.61		0.945
PBD2	81.45	149.28	256.88	319.89	390.60	407.02	380.30	490.17		0.950
PBE1	81.01	107.76	196.21	302.90	365.03	404.25	378.67	441.03		0.981
PBE2	0.00	1.52	5.59	8.05	40.06	56.56	59.39	58.56		0.887
PBF1	71.53	112.26	190.45	226.89	259.90	366.60	431.42	420.23		0.943
PBG1	104.93	174.55	232.34	392.37	425.10	397.85	396.42	463.13		0.906
									Average	**0.845**

pipeline was 1.9-fold faster than the TPP/ABACUS pipeline (Figure 2.5). Such an improvement was expected given that Morpheus completes its searches on average 1.3 to 4.6 times faster than most other search engines available (Wenger and Coon, 2013). Given its simplicity of use, speed, and open source nature, MSpC combined with Morpheus is clearly advantageous over other PSM-based LFQ approaches currently available. Moreover, by being open source, MSpC should allow others to

extend its utility and to serve as a platform for integrating additional open source LFQ approaches into the Morpheus pipeline.

2.3. Methods

2.3.1. Sample Preparation

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. 20S proteasomes were obtained as described previously (Book et al., 2010) from *Arabidopsis thaliana* Col-0 ecotype seedlings in which the 20S proteasome subunit PAG1 (α7) was genetically replaced with a FLAG-tagged variant, with the minor modification of switching to a more stable HEPES buffer during purification. The FLAG peptide used for elution was removed by filtering through an Amicon Ultra 4 10K filter with the elution buffer also exchanged into 8 M urea. Total protein was quantified by the bicinchoninic acid protein assay (Thermo Scientific) using bovine serum albumin as the standard. Approximately 70 μg of proteasomes were digested overnight at 37°C using a 1:30 trypsin/sample ratio. Peptides were acidified to a final concentration of 1% TFA, desalted on a Waters C18 Sep-Pak containing 50 mg sorbent material, and lyophilized. Total *E coli* lysates were obtained from Bio-Rad (Cat. 163-2110) with 200 μg digested as above. Both proteasome and *E. coli* peptides were dissolved in 5% acetonitrile, 95% water, and 0.1 % formic acid. Each MS analysis, performed
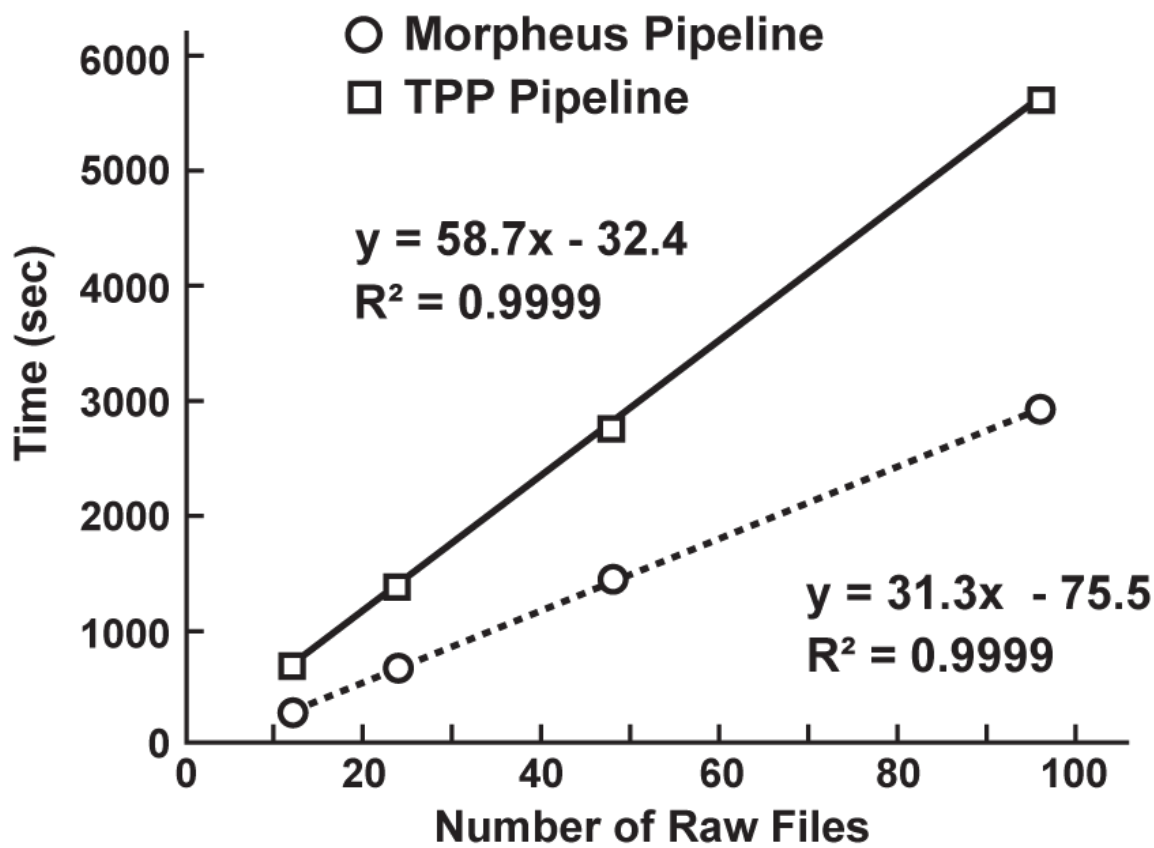
Figure 2.5: MSpC combined with Morpheus works faster than TPP combined with ABACUS. Speed comparisons were performed for 12, 24, 48, and 96 raw MS/MS files generated with the 20S proteasome/*E. coli* lysate samples analyzed in 2.4. On average, Morpheus/MSpC finished the calculations 1.9 times faster than TPP/ABACUS over a ~10-fold range of dataset size.

in triplicae, used 5 µL volumes prepared with 3, 2, 1.5, 1, 0.5, 0.25, 0.1, or 0.05 µg of digested proteasomes mixed with 0.5 µg of digested E. c.oli proteins. This mixtures reflected proteasome/*E. coli* ratios of ~0.091, 0.167, 0.333, 0.500, 0.667, 0.750 0.800, and 0.857, respectively.

2.3.2.   Liquid Chromatography and High-Resolution Mass Spectrometry

Samples were analyzed by ultra-high performance liquid chromatography (UPLC) (nanoAcquity, Waters Corporation) connected online to an electrospray ionization LTQ Velos Orbitrap mass spectrometer (Thermo Scientific). Separation employed a 100 x 365 µm fused silica capillary micro-column packed with 20 cm of 1.7 µm-diameter, 130-Å pore size, C18 beads (Waters BEH), with an emitter tip pulled to approximately 1 µm using a laser puller (Sutter Instruments). Peptides were loaded at a flow-rate of 400 nL/min for 30 min and then eluted over 120 min at a flow-rate of 300 nL/min with a 2% to 30% acetonitrile gradient in 0.1% formic acid. Full-mass scans were performed in the FT Orbitrap with a mass range of 300-1500 m/z at a resolution of 60,000, followed by ten MS/MS high energy C-trap dissociation scans of the ten highest intensity parent ions at 42% normalized collision energy and 7,500 resolution, with a mass range starting at 100 m/z. Dynamic exclusion was enabled with a repeat count of two over the duration of 30 sec and an exclusion

window of 120 sec.

2.3.3.   Data Processing

Protein identifications were determined using the Morpheus search engine (Wenger and Coon, 2013). Raw data was searched with the Thermo module of Morpheus revision 151 downloaded and compiled from source code available at `http://sourceforge.net/projects/morpheus-ms/` using Microsoft Visual Studio 2013 professional edition. The following parameters were used to search all databases: unknown precursor charge states - +2, +3, +4; maximum number of MS/MS peaks = 400; assign charge states - enabled; de-isotope - disabled; generate target decoy database on the fly; protease trypsin (no proline rule); maximum missed cleavages = 2; initiator methionine behavior - variable; fixed modification of carbamidomethylation of cysteine; variable modification of oxidation of methionine; maximum variable modification isoforms per peptide = 1024; precursor mass tolerance = $\pm$ 2.1 Da monoisotopic (recommended parameters to account for neutral loss); precursor monoisotopic peak correction - disabled; product mass tolerance = $\pm$ 0.025 Da monoisotopic; consider modified forms as unique peptides - false; maximum threads = 8; minimize memory usage - false. MSpC quantification of Universal Proteome Standard 2 (UPS2) protein sequences exploited the MS/MS analysis of

individual USP2 proteins (Sigma-Aldrich) mixed at various concentrations with egg or embryo extracts from Xenopus laevis available in the PRoteomics IDEntifications (PRIDE) repository (Vizcaino et al., 2013) using identifier - PXD000902 and the available proteomics database - pita_v1.71.protein.name.fa. For our analysis, both the raw MS/MS data and resulting FASTA files for the USP2 and egg and embryo proteomes were obtained from PRIDE. The database used for searching the MS/MS data of *Arabidopsis* proteasomes spiked into total *E. coli* peptides was generated by combining Uniprot K12 *E. coli* reference proteome UP000000625 with a common contaminant database, and then mixing the merged dataset with FASTA sequences for all proteoforms of all known proteasome subunits and associated proteins (Book et al., 2010) obtained from the TAIR10_pep_20101214 FASTA database available within The *Arabidopsis* Information Resource (TAIR) version 10. All FASTA files are available for download in the Supporting Information. The datasets were analyzed by MSpC with a 1% PSM false discovery rate (FDR) and a 1% protein group FDR to determine NSAF, dNSAF, and uNSAF values for each protein group. Two separate analyses were performed in which one unique, or alternatively two unique peptides were required to quantify a protein group. NSAF values were calculated according to formulas 1a, 2a, and 3a from Figure 1 of Zhang et al. (Zhang et al., 2010).

2.3.4. Isoform Incorporation Rates

For the individual subunit analysis of the 20S proteasome the isoform incorporation rates were treated as follows. Given that each proteasome subunit should be incorporated at equal stoichiometry within the 20S particle, we then tested whether the Morpheus/MSpC pipeline could calculate the relative abundance of each subunit and the distribution of isoforms. Here, we divided the dNSAF values for each subunit/isoform by the total number of dNSAF values for the entire complex across all eight total proteasome/*E. coli* lysate ratios tested. This averaged value provided a concentration-independent ratio for the incorporation of each subunit/isoform. We then normalized these values based on a 1/14 stoichiometry of each subunit within the complex to calculate the estimated occupancy versus the expected occupancy of each subunit.

2.3.5. Speed and Accuracy Comparisons of Morpheus/MSpC to the TPP/ABA-CUS Pipeline

The speed and accuracy of MSpC combined with Morpheus was compared to the next most comparable open source software suite for calculating NSAF values; i.e., TPP (Deutsch et al., 2010) combined with ABACUS (Fermin et al., 2011), using the proteasome spike-in experiment files as input. The .raw files were converted to

.mzmL files by TPP Build 201411201551-6764 and then searched using the multi-threaded X!Tandem MS/MS search engine (Craig and Beavis, 2004) with the search parameters adjusted as close as possible to that used for the Morpheus searches. A decoy database was generated using the TPP tool DecoyFASTA for use with X!Tandem. Relevant X!Tandem parameters are listed here: parent monoisotopic mass error = $\pm$ 2.1 Da, fragment mass error = $\pm$ 0.025 Da, fixed modifications of carbamidomethylation (57.021464) on cysteine, and variable modification of oxidation (15.994915) on methionine, fully tryptic cleavages, missed cleavage sites = 2 maximum, no refinement and 8 threads. The configuration file used and the test datasets can be found in the Supporting Information. The data were analyzed in the TPP using Peptide Prophet (Keller et al., 2002). Relevant settings are listed: minimum probability = 0.05; minimum peptide length = 7; accurate mass, and nonparametric decoy database to pin down false discovery rate; ignore +1 charged spectra; and run Protein Prophet (Nesvizhskii et al., 2003) after Peptide Prophet. Once completed, all pepxml data from Peptide Prophet contained in a single folder was combined using the command line version of Protein Prophet from the TPP binaries with the following command: ProteinProphet.exe *.pep.xml interact-COMBINED.prot.xml. This post-analysis aggregation was required for running the spectral counting program ABACUS. Here, we note that there are no graphic user interfaces to perform this

post analysis aggregation, which makes this portion of the data analysis more diffi-
cult for those unfamiliar with setting up and running programs from the command
line. The combined data was analyzed by ABACUS with the following parameters:
best peptide probability = 0.99; minimum peptide probability = 0.99; experimental
peptide probability = 0; and combined file probability = 0.99 to most accurately
match a 1% FDR stringency settings in MSpC. dNSAF values were compared in
Microsoft Excel using the CORREL function and squaring the result. Additional
timing tests were performed with a subset of the calibration curve data (ratios 0.091,
0.167, 0.333, and 0.500 in triplicate corresponding to 12 .raw files) by increasing file
input to 24 (2x), 48 (4x), and 96 (8x) .raw files to determine the time dependence on
input size between both pipelines tested (Morpheus/MSpC versus TPP/ABACUS).
The timing tests and all data analyses were performed on a computer running
Windows 7 Ultimate, with 16 GB of random access memory, and an Intel Core
i7-2700k with hyper-threading turned on for eight logical cores.

2.4.   Discussion on Requiring Two Unique Peptides to Quantify a Protein

Occasionally, some researchers may want to use a more stringent criterion for quan-
tification such as requiring a protein to have more than one unique identifying
peptide. To see how this might affect our data analysis, we re-analyzed our results

shown in Figure 2.3, this time requiring two unique peptides to quantify an individual protein. The results point to a very small increase in linearity observed in the average plots; however, there is a slight decrease in linearity for the egg sample (0.886 to 0.865) and a larger decrease in linearity for the individual UPS2 protein plot for the embryo sample (0.827 to 0.723). The decrease in linearity in the embryo sample is due to the analysis removing a low abundance UPS2 protein (O00762ups) identified with only one unique peptide. While some have suggested that requiring more than one unique peptide to identify a protein is an ideal approach, requiring two peptides for identifications in database searches reduces the number of protein identifications in the target database more than those in the decoy database and results in increased false discovery rates . While we recognize that researchers may want to implement more stringent requirements than what is typically used in database searching to quantify a set of proteins, there are two cases where requiring two unique peptides may not be ideal in a quantitative analysis. Firstly, low abundance proteins that have few PSMs might be identified by only a single peptide and thus be erroneously thrown out of the analysis. Secondly, there may be only one unique peptide that can differentiate between families of homologous proteins. In this second case, requiring two unique peptides would remove these homologous proteins from the MSpC analysis, even if they had a large number of

PSMs. Because of these reasons and because of the decreased linearity observed when requiring proteins to have two unique peptides (Figure 2.3) as compared to one unique peptide (Figure 2.2), we suggest caution in requiring more than one unique peptide per protein.

2.5. Tutorial

Morpheus Spectral Counter (MSpC) Tutorial

Requirements

- 64 bit Windows Installation: For Example Microsoft Windows 7 64 bit, Windows 8 and 8.1 64bit, or Windows 10 64bit

    - This is due to requirements from vendor libraries to process Thermo .raw files directly in this tutorial

- Download and install proteowizard (http://proteowizard.sourceforge.net/downloads.shtml)

    - Installation of proteowizard correctly installs the latest Thermo Vendor .dll's to read .raw files directly

- Download pre-compiled binaries of revision 151 of the Morpheus Mass Spectrometry Search Engine (Thermo Version) (https://github.com/dcgemperline/Morpheus_SpC/releases/download/v1.0/Morpheus.Binaries.revision.151.zip) , referred from here on out as just Morpheus

- Download and install MSpC (https://github.com/dcgemperline/Morpheus_SpC/releases/download/v1.0/MSpC_v1.0.zip)
    - MSpC depends on the latest .NET Runtime that will be installed with MSpC installer if it is not already installed

- Download the data files from the PRIDE proteomics data repository using the following ID - PXD003002 Data Files (http://www.ebi.ac.uk/pride/archive/)

Data Analysis Starting from .raw files

1. Once all the requisite software is installed and the Data Files are downloaded you are ready to begin.
2. Unzip **Morpheus Binaries revision 151.zip** into your desired location and start **Morpheus (Thermo).exe** from the Morpheus (Thermo) folder.
3. Once Morpheus starts, verify at the top you are using **Morpheus (Thermo) revision 151**.
4. Add the following .raw files to Morpheus from the Raw Files folder contained in the Data Files folder (multiple files can be selected at once).
    - 04_28_15_A_rep1_c1.raw
    - 04_28_15_A_rep2_c1.raw

- 04_28_15_A_rep3_c1.raw
- 04_28_15_B_rep1_c1.raw
- 04_28_15_B_rep2_c1.raw
- 04_28_15_B_rep3_c1.raw
- 04_28_15_C_rep1_c1.raw
- 04_28_15_C_rep2_c1.raw
- 04_28_15_C_rep3_c1.raw
- 04_28_15_D_rep1_c1.raw
- 04_28_15_D_rep2_c1.raw
- 04_28_15_D_rep3_c1.raw

5. Side Note on Fractionation: If you would like Morpheus, and thus MSpC to output summaries for a set of data by adding up all of the spectra identified in that dataset(as the case may be for fractionation data), place the files you would like to be summarized in seperate folders, such as SampleSet1, Sampleset2, and Morphues will generate protein_groups.tsv and PSMs.tsv specifically for each sampleset. Later in MSpC after selecting the summary.tsv file these will show up as SampleSet1*, SampleSet2*, etc.

6. Browse for the fasta file contained in the Data Files Folder **uniprot_k12_e_coli_contams_plus_proteasome_and_interactors.fasta**.

7. Verify that Create Target-Decoy Database On The Fly is checked.

8. Browse for an Output Folder. In this tutorial select browse, leave the default Desktop location highlighted and press Make a New Folder and rename it Morpheus Analysis. Press OK. You should now have a folder on the desktop called Morpheus Analysis, and Morpheus should say you are outputting data to the following output folder, where *NAME* is your username on your machine.

   - C:\Users*NAME*\Desktop\Morpheus Analysis

9. Set the Maximum threads to 2, for a dual core processor, 4 for a quad core processor, 8 for a quad core processor with hyperthreading, and 8 for an 8 core processor. These are reasonable values that will give you decent performance with Morpheus.

10. Press Search, and the progress bar will indicate search progress for each raw file. On an Intel Core i7 2700K with Morpheus set to use 8 threads, this takes approximately 4.5 minutes.

11. Close Morpheus if so desired.

12. On completion of the search open MSpC.

13. The indicators in Summary Directory Contains Valid Input should be red.

14. Press the Select the Summary File button and navigate to the summary.tsv file contained in the Morpheus Analysis folder on the Desktop.

    - C:\Users*NAME*\Desktop\Morpheus Analysis\summary.tsv

15. The indicators in Summary Directory Contains Valid Input should now read green indicating that the Morpheus Analysis folder contains all the necessary files from Morpheus to run. (The files requried to be in the same directory as summary.tsv are protein_groups.tsv and PSMs.tsv. This will automatically occur if you output all of your Morpheus output into a single output folder.)

16. (Optionally) Select the provided whitelist file contained in the Data Files folder that you downloaded to simplify the output.

17. Select your desired PSM FDR and Protein FDR (although the default options of 1% are good default options).

18. Check or uncheck boxes of the desired calculations (NSAF, dNSAF, uNSAF)

19. (Optionally) Select an Output Directory, otherwise MSpC defaults to the current directory of summary.tsv as an output directory.

    - Let's press Select Output Directory then Make a New Folder here called MSpC Output, press OK.

20. Press Summarize Data, and the progress bar will indicate the progress of MSpC. This should take ~ 30 seconds or less

21. The MSpC Output folder will then contain NSAF summaries for each individual .raw file, as well as a summary for all .raw files analyzed. These files are in the tab delimited output format .tsv

Data Analysis Starting from previously searched Morpheus Data

1. It is strongly suggested, that you start with the Data Analysis from .raw files tutorial first, and then refer back to the tutorial for previously searched Morpheus data if necessary

2. Open MSpC

3. Select a Morpheus summary.tsv file in a directory that contains the entire output from Morpheus.

4. If all output is there, the indicators in Summary Directory Contains Valid Input should now read green indicating that the folder contains all the necessary files from Morpheus to run MSpC.

5. (Optionally) Select the provided whitelist file contained in the Data Files folder that you downloaded to simplify the output.

6. Select your desired PSM FDR and Protein FDR (although the default options of 1% are good default options).

7. Check or uncheck boxes of the desired calculations (NSAF, dNSAF, uNSAF)

8. (Optionally) Select an Output Directory, otherwise MSpC defaults to the current directory of summary.tsv as an output directory.

    - Let's press Select Output Directory then Make a New Folder here called MSpC Output, press OK.

9. Press Summarize Data, and the progress bar will indicate the progress of MSpC. This should take ~ 30 seconds or less

10. The MSpC Output folder will then contain NSAF summaries for each individual .raw file, as well as a summary for all .raw files analyzed. These files are in the tab delimited output format .tsv

2.6. Acknowledgements

2.7. References

Book, A.J., Gladman, N.P., Lee, S.S., Scalf, M., Smith, L.M., and Vierstra, R.D. (2010). Affinity purification of the arabidopsis 26s proteasome reveals a diverse array of plant proteolytic complexes. J. Biol. Chem. **285**(33):25554–25569.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed maxlfq. Mol. Cell. Proteomics **13**(9):2513–2526.

Craig, R. and Beavis, R.C. (2004). Tandem: matching proteins with tandem mass spectra. Bioinformatics **20**(9):1466–1467.

Deutsch, E.W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J.K., Martin, D.B., Nesvizhskii, A.I., and Aebersold, R. (2010). A guided tour of the trans-proteomic pipeline. Proteomics **10**(6):1150–1159.

Fermin, D., Basrur, V., Yocum, A.K., and Nesvizhskii, A.I. (2011). Abacus: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. Proteomics **11**(7):1340–1345.

Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P. (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem ms. Proc. Natl. Acad. Sci. USA **100**(12):6940–6945.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by ms/ms and database search. Anal. Chem. **74**(20):5383–5392.

Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. **75**(17):4646–4658.

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics **1**(5):376–386.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D.J. (2004). Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics **3**(12):1154–1169.

Russell, J.D., Scalf, M., Book, A.J., Ladror, D.T., Vierstra, R.D., Smith, L.M., and Coon, J.J. (2013). Characterization and quantification of intact 26s proteasome proteins by real-time measurement of intrinsic fluorescence prior to top-down mass spectrometry. PLoS One **8**(3):e58157.

Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms. Anal. Chem. **75**(8):1895–1904.

Tu, C., Li, J., Sheng, Q., Zhang, M., and Qu, J. (2014). Systematic assessment

of survey scan and ms2-based abundance strategies for label-free quantitative proteomics using high-resolution ms data. J. Proteome Res. **13**(4):2069–2079.

Vizcaino, J.A., Cote, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Perez-Riverol, Y., Reisinger, F., Rios, D., Wang, R., and Hermjakob, H. (2013). The proteomics identifications (pride) database and associated tools: status in 2013. Nucleic Acids Res. **41**(Database issue):D1063–9. Vizcaino, Juan Antonio Cote, Richard G Csordas, Attila Dianes, Jose A Fabregat, Antonio Foster, Joseph M Griss, Johannes Alpi, Emanuele Birim, Melih Contell, Javier O'Kelly, Gavin Schoenegger, Andreas Ovelleiro, David Perez-Riverol, Yasset Reisinger, Florian Rios, Daniel Wang, Rui Hermjakob, Henning eng BB/I024204/1/Biotechnology and Biological Sciences Research Council/United Kingdom WT085949MA/Wellcome Trust/United Kingdom Research Support, Non-U.S. Gov't England 2012/12/04 06:00 Nucleic Acids Res. 2013 Jan;41(Database issue):D1063-9. doi: 10.1093/nar/gks1262. Epub 2012 Nov 29.

Wenger, C.D. and Coon, J.J. (2013). A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. J. Proteome Res. **12**(3):1377–1386.

Wong, J.W. and Cagney, G. (2010). An overview of label-free quantitation methods in proteomics by mass spectrometry. Methods Mol. Biol. **604**:273–283.

Yang, P., Fu, H., Walker, J., Papa, C.M., Smalle, J., Ju, Y.M., and Vierstra, R.D. (2004). Purification of the arabidopsis 26s proteasome: biochemical and molecular analyses revealed the presence of multiple isoforms. J. Biol. Chem. **279**(8):6401–6413.

Zhang, B., VerBerkmoes, N.C., Langston, M.A., Uberbacher, E., Hettich, R.L., and Samatova, N.F. (2006). Detecting differential and correlated protein expression in label-free shotgun proteomics. J. Proteome Res. **5**(11):2909–2918.

Zhang, Y., Wen, Z., Washburn, M.P., and Florens, L. (2010). Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. Anal. Chem. **82**(6):2272–2281.

Zybailov, B., Mosley, A.L., Sardiu, M.E., Coleman, M.K., Florens, L., and Washburn, M.P. (2006). Statistical analysis of membrane proteome expression changes in saccharomyces cerevisiae. J. Proteome Res. **5**(9):2339–2347.

COLOPHON

This document was typeset with LaTeX. It is based on the University of Wisconsin dissertation template created by William C. Benton (available at `https://github.com/willb/wi-thesis-template`).